



ASOCIACION ARGENTINA  
DE ECONOMIA POLITICA

ANALES | ASOCIACION ARGENTINA DE ECONOMIA POLITICA

# LIII Reunión Anual

Noviembre de 2018

ISSN 1852-0022

ISBN 978-987-28590-6-0

Aprendizaje automático y métodos hedónicos en  
el mercado de autos usados en línea de Argentina

**Gutiérrez Emiliano Martín**

**Larrosa, Juan Manuel Ceferino**

**Delbianco Fernando**

**Uriarte, Juan Ignacio**

**Ramírez Muñoz de Toro Gonzalo Román**

# **Aprendizaje automático y métodos hedónicos en el mercado de autos usados en línea de Argentina**

Emiliano M. Gutiérrez (1,2), Juan M.C. Larrosa (1,2), Fernando Delbianco (1,3), Juan I. Uriarte (5) y Gonzalo Ramírez Muñoz de Toro (4,5)

1. Departamento de Economía, Universidad Nacional de Sur
2. Instituto de Investigaciones Económicas y Sociales del Sur (IIESS)
3. Instituto de Matemática de Bahía Blanca (INMABB)
4. Departamento de Derecho, Universidad Nacional del Sur
5. Hyperia

## **Resumen**

El presente trabajo utiliza métodos de machine learning para reducir el número de características determinantes del precio de automóviles usados en Argentina. Nos basamos en la especificación de un modelo de precios hedónicos en línea para los automóviles usados (Ramírez Muñoz del Toro et al ,2017). Específicamente aplicamos el método Least Absolute Shrinkage and Selection Operator (LASSO) y el Classification and Regression Tree (CART) junto a una estimación más tradicional de un modelo hedónico. Los datos fueron obtenidos de un sitio en línea. Mediante el uso de estas técnicas nos es posible realizar una selección de variables relevantes y explorar posibles no linealidades, que complementan el análisis de regresión tradicional.

## **1. Introducción.**

Desde la perspectiva de la teoría microeconómica clásica, el equilibrio dentro de un mercado para un determinado bien o servicio se obtiene a partir de la intersección de la oferta y la demanda. De esta manera el nivel óptimo es alcanzado de manera tal que tanto la valoración de los consumidores como la de los productores, es exactamente la misma. No obstante esta asignación considerada de eficiencia, adolece la existencia de grandes supuestos, siendo uno de ellos la presencia de información perfecta por parte de todos los agentes.

En este sentido Akerlof (1970), observa que cuando se relaja este último postulado, analizando el caso de los autos usados, se genera la existencia de asimetrías en la información: el vendedor tiene mayor información respecto del consumidor respecto de la calidad del automóvil.

Lancaster (1966) supone que los consumidores obtienen utilidad de los atributos de bienes. Suponiendo que un automóvil es el único bien involucrado en el consumo de actividad de conducción, éste produce un vector fijo de atributos y su nivel de actividad es un escalar asociado con el vector (relación que podría ser lineal).

Resulta entonces de interés analizar cuál sería la valoración e importancia por parte de los consumidores de los diversas particularidades que puede poseer un coche usado. Una herramienta a la cual puede apelarse para ilustrar este fenómeno viene dada por los modelos de precios hedónicos, los cuales pueden ser útiles para determinar el valor intrínseco de cada atributo, así como para predecir los precios de transacción. Un bien es comprendido como un conjunto de características reflejadas en los precios implícitos que pueden ser derivados conteniendo diferentes combinaciones de cualidades específicas.

A su vez, la aparición del Big Data, y la aplicación de técnicas de Machine Learning (ML) al estudio de datos, otorga a la ciencia económica una gran oportunidad hacer uso de dichas herramientas para complementar el análisis tradicional (Varian, 2014, Ng, 2017).

Consecuentemente este trabajo se propone evaluar un modelo de precios hedónicos para automóviles usados a partir de la información disponible en línea dentro de Argentina. De este modo, el presente artículo se estructurará de la siguiente forma: en la primera parte se revisarán aquellos antecedentes considerados de relevancia, la sección II expondrá los datos y metodología utilizada, mientras que en la sección III, se exhibirán los resultados. Finalmente en la sección IV se desarrollarán las principales conclusiones.

## **2. Modelos de precios hedónicos y Machine Learning (ML): antecedentes conceptuales y metodológicos.**

Un estudio de precios hedónicos tiene como finalidad demostrar la relaciones entre las diversas características que hacen parte de un bien compuesto y su precio, como así también describir las preferencias individuales que impulsan a los agentes a adoptar sus decisiones de compra (Gilbert,2013).

Este tipo de análisis hallan su razón de ser a partir del aporte de Court (1939), quien analiza diversos índices de precios hedónicos para piezas automotrices, partiendo de la base de que las características de un bien afectaban a su valoración —su “utilidad y deseabilidad” en términos del mismo Court (1939: 107).

Siguiendo esta línea, debe resaltarse el aporte de Griliches (1961), quien recurre a los precios de automóviles entre 1957 y 1960 provenientes de Estados Unidos. A su vez, Cowling y Cubbin (1972), postulan la existencia de modelos hedónicos para la creación índices de precios de automóviles en el Reino Unido a fin de realizar ajustes cualitativos tanto por la variedad como por su evolución en el tiempo.

Un número considerable de métodos han sido puesto a discusión a fin de evaluar el impacto de los atributos al momento de que los consumidores tomen decisiones (Gilbert, 2013). Rosen (1974), describe a los automóviles por las  $n$  características medibles (por ejemplo, consumo de combustible, tamaño del vehículo, motor y tecnología, entre muchas otras) y un vector  $Z = (Z_1, Z_2, \dots, Z_n)$  con  $Z_n$  midiendo la cantidad de las  $i$ -ésimas características.

De esta manera, una primera aproximación de un modelo típico de precios hedónicos puede ser representada como

$$Y_i = f(Z_i) + \varepsilon_i$$

Así las cosas, debe resaltarse el surgimiento de una línea de investigación que dio lugar a la aparición de modelos de precios hedónicos vinculados con el sector automotriz a causa del establecimiento de las regulaciones sobre el consumo de combustible de los vehículos vendidos en Estados Unidos a mediados de la década del '70 y principios de los '80<sup>1</sup>, lo cual ha dio lugar al advenimiento de contribuciones que merecen ser reconocidas. De esta forma Boyd y Mellman (1980), resultan precursores en estudiar los efectos del combustible, considerando autos usados

---

<sup>1</sup> Los estándares promovidos por el Corporate Average Fuel Economy (CAFE), fueron regulaciones sancionadas en los Estados Unidos, en 1975. Las mismas establecían ciertos niveles de consumo de combustible de los autos dependiendo el año de venta, en caso de incumplimiento en el diseño, la empresa manufacturera debía afrontar una penalización monetaria

y nuevos como bienes sustitutos. Dichos autores realizan un análisis mediante el cual estiman un modelo logit, en que, además de plantear al consumo en gasolina como una variable explicativa, recurren a otros atributos propios del vehículo (maniobrabilidad, reparaciones efectuadas, ruido, aceleración, espacio interior).

A pesar de este aporte, debe resaltarse la ausencia de una elevada cantidad estudios de precios hedónicos que consideren al precio del combustible como un determinante de importancia al momento de adquirir un automóvil. No obstante Kiso (2015) propone una estimación de tipo log-log, evaluando mediante OLS y mínimos cuadrados generalizados (GLS), para el mercado estadounidense, considerando además del consumo en gasolina otras variables como el peso, tipo de transmisión, potencia, tipo de tracción y tipo de vehículo. A su vez Saltee et al. (2016) plantean si los consumidores valoran realmente el costo del combustible al momento de adquirir un automóvil. Con datos provenientes de las transacciones realizadas en Estados Unidos entre 1993 y 2008, concluyen que frente a un incremento esperado en el costo del combustible, se reducirá el precio de venta. En esta misma dirección Fan y Rubin (2010), utilizan la información de aquellos autos registrados en Maine (Estados Unidos), para observar la disposición a pagar respecto del consumo de combustible de un coche nuevo.

En otra línea Baltas y Saridakis (2009) destacan la existencia de un mercado de vehículos segmentado, de forma tal que evalúan mediante tres modelos de precios hedónicos log-lin el impacto dado por la marca, apariencia y atributos propios del auto. Dicho modelo es denominado B&S (en clara alusión a los autores del trabajo), por Kihm y Vance (2016) quienes comparan este último trabajo con estimaciones propias y resaltando que el principal determinante que relaciona el precio de un modelo nuevo y uno usado es el consumo de combustible, siendo los que menos consumen aquellos que resultan un valor mayor. De igual modo, González et al. (2015) analizan mediante un modelo de datos envolventes (DEA), para el mercado de autos español, procurando demostrar la fuerte segmentación en la cual se halla inmerso el mercado de automóviles. De esta forma realizan una estimación considerando el descuento que puede realizar la firma respecto del precio de lista y con las mejoras adicionales que pueden incorporarse a un automóvil.

Las estrategias en la comercialización también son relevadas por Dvir y Strasser (2018) quienes estudian el mercado de automóviles para la Unión Europea durante 1993-2011 a fin de testear la convergencia en el valor hedónico de venta, encontrando que si bien parte de este último era exógeno a las firmas, observan también que la presencia de estrategias llevadas a cabo por las compañías automotrices contribuyen a la dispersión en los precios.

Por otra parte, Desmet y Hendoui (2000) analizan el mercado inglés de pequeños autos en el que consideran 48 determinantes del precio de los automóviles, recurriendo a una estimación

lineal típica de Mínimos Cuadrados Ordinarios (OLS), como así también, de manera innovadora, se utilizan técnicas de estimación de redes neuronales (RNA).

De igual forma Prieto et al. (2015) examinan el mercado de autos usados francés en línea considerando cuatro modelos. Sin embargo, dada la pretensión de diferenciar entre vendedores expertos y aquellos ocasionales, recurren a dos variables instrumentales: kilómetros redondos y tiempo de posteo. Desde el punto de vista de estos autores, un anuncio cuyo tiempo de publicación es corto, como así también la indicación de un kilometraje de manera aproximada, pueden considerarse como señales de que se trata de un comerciante experto, a diferencia de un vendedor ocasional.

Por su parte, Erdem y Sentürk (2009), consideran un modelo hedónico de precios en el que evalúan el mercado de autos usados de Turquía, recurriendo a información en línea, realizando estimaciones log-log, lin-log y transformación de Box-cox. Este aporte es posteriormente retomado por Akay et al. (2018), quienes para la misma zona geográfica evalúan el precio de los automóviles de segunda mano mediante OLS, mínimos cuadrados recortados (LTS), mínimas desviaciones absolutas (LAD) y estimaciones de tipo S, MM y M.

Para el caso concreto del ámbito nacional, Ramírez Muñoz de Toro et al. (2017), analizan el mercado de autos argentinos de segunda mano, estimando diversas especificaciones de modelos OLS de tipo log-lin, a partir de la recolección de datos de un sitio web, durante el año 2015.

En el campo de Machine Learning (ML), un valioso aporte es realizado por Pudaruth (2014) quien recurre a técnicas de este tipo para extraer precios de coches usados, concretamente para Mauritania. Dicho trabajo es retomado por Knoor y Jan (2017) quienes estiman el precio de los autos tanto nuevos como usados para Pakistán recurriendo a OLS, y obteniendo la información a partir de un sitio web dedicado exclusivamente a la venta automotriz.

Otro aporte, desde el campo de ML, es realizado por Oladunni, et al. (2017), donde el objetivo de esta investigación consiste en plantear un modelo hedónico de precios, considerando el mercado inmobiliario de Washington D. C., durante la burbuja y posterior crisis subprime. Como métodos de estimación recurren a LASSO (Least Absolute Shrinkage and Selection Operator), como así también a Random Forest.

Por su parte, Yang (2018), realiza una extensa y valiosa investigación en la que aborda diversos métodos desde la ML para estimar modelos de precios hedónicos considerando el valor de las propiedades afectadas por el fracking, para el estado de Pennsylvania (Estados Unidos) durante el período 2000-2012. Recurriendo a la extracción de publicaciones provenientes de un sitio web referentes a las características propias de la vivienda en venta, en simultáneo con la información censal, como así también acudiendo al procesamiento de imágenes de las mismas

para otorgarle un puntaje basado en criterios estéticos. Mediante estimaciones de tipo LASSO, OLS, Random Forest, Bagging y Boosting, analiza entre 172 y 77 variables, dependiendo el método utilizado.

Asimismo resulta menester señalar que si bien consta una extensa literatura respecto a la existencia de modelos hedónicos al momento de explicar el precio final de un vehículo, se debe resaltar la ausencia de una teoría concreta que determine la forma funcional apropiada al momento de realizar estas estimaciones (Cropper et. al., 1988).

### 3. Metodología y datos

Los datos utilizados en este trabajo provienen del sitio Mercado Libre (www.mercadolibre.com.ar) en el periodo comprendido desde marzo hasta julio del 2015, donde la cantidad de observaciones recopilada es de 920, correspondiendo cada una de ellas a un aviso de venta de un auto de segundo mano.

A diferencia de gran parte de las investigaciones anteriores respecto a la realización de modelos hedónicos de precios para autos usados, se recurrirá a una estimación de tipo LASSO (Least Absolute Shrinkage and Selection Operator) como así también se hará uso de CART (Classification and Regression Tree).

De esta forma, los atributos a ser utilizados para ambas estimaciones se detallan en la tabla 1. Dichas variables son las seleccionadas por Ramírez Muñoz de Toro et al. (2017), de manera tal que se recurre a una consideración de atributos generales, técnicos y de equipamiento<sup>2</sup>.

**Tabla 1. Características de los vehículos analizadas**

---

Airbag pasajero (binaria) / v19	Doble tracción (binaria) / v51
Silla de bebé (binaria) / v38	DVD (binaria) / v42
Airbag conductor (binaria) / v18	Entrada auxiliar (binaria) / v6
Airbag de cortina (binaria) / v48	Entrada USB (binaria) / v28
Airbags laterales (binaria) / v33	Espejos eléctricos (binaria) / v14
Aire acondicionado (binaria) / v8	Faros antiniebla delanteros (binaria) / v36
Alarma (binaria) / v20	Faros antiniebla traseros (binaria) / v39
Alarma de luces encendidas (binaria) / v9	Faros de xenón (binaria) / v37

---

<sup>2</sup> Para mayor información de los criterios utilizados para la selección de variables, se sugiere recurrir directamente a Ramírez Muñoz de Toro y otros (2017)

Alta gama (binaria) / Alta gama	Faros regulables desde el interior (binaria) / v31
Am/Fm (binaria) / v5	Frenos ABS (binaria) / v17
Año (año de construcción) / A	GPS (binaria) / v52
Apertura remota de maletero (binaria) / v29	Importado (binaria) / Importado
Apoyacabeza asientos traseros (binaria) / v21	Inmovilizador de motor (binaria) / v22
Argentina (binaria) / Argentina	Kms (cantidad de kilómetros del vehículo) / KM
Asiento conductor regulable(binaria) / v11	Limpia-lavaluneta (binaria) / v1
Asiento trasero rebatible (binaria) / v10	Llantas de aleación (binaria) / v2
Asientos eléctricos (binaria) / v49	Marca (categórica para 33 marcas) / M
Blindado (binaria) / v50	MERCOSUR (binaria) / Mercosur
Bluetooth (binaria) / v24	MP3 (binaria) / v7
Caja de CD (binaria) / v25	Paragolpes pintados (binaria) / v3
Cargador de CD (binaria) / v26	Pasacasete (binaria) / v53
Cierre centralizado de puertas (binaria) / v12	Pick-up (binaria) / PickUp
Climatizador automático (binaria) / v30	Potencia (HP) / Potencia
Color (categoría numérica arbitraria) / C	Puertas (cantidad de puertas) / PTA
Comando satelital para estéreo (binaria) / v41	Repartidor electrónico de fuerza de frenado (binaria) /v40
Combustible Diésel / COMBDIE	Sensor de estacionamiento (binaria) / v44
Combustible GNC / COMBGNC	Sensor de lluvia (binaria) / v45
Combustible Nafta / COMBNAF	Sensor de luz (binaria) / v46
Computadora de abordó (binaria) / v13	Tapizado de cuero (binaria) / v47
Control de estabilidad (binaria) / v35	Tarjeta SD (binaria) / v27
Control de tracción (binaria) / v34	Techo corredizo (binaria) / v32
Control de velocidad de cruceo (binaria) / v43	Tercera luz de freno (binaria) / v16
Cristales eléctricos (binaria) / v15	Transmisión (automática o manual) / Trans
Década (categórica para 3 décadas) / década	Ubicación del vendedor (distancia respecto a CABA) / U
Dirección (automática, secuencial) / D	Único dueño (binaria) / UD
	Utilitario (binaria) / Utilitario
	Vidrios polarizados (binaria) / v4

Fuente: Ramírez Muñoz de Toro et al. (2017)

### 3.1 LASSO (Least Absolute Shrinkage and Selection Operator)

A partir de la clásica regresión multivariada, en la que una serie de un número  $p$  de variables explicativas  $X$ , con  $Y$  como variable dependiente y un intercepto, Se desea predecir la variable de interés, concretamente el precio informado del vehículo para este caso. Además se supone que todas las regresoras han sido estandarizadas con media cero y varianza uno, puede incorporarse un término de penalización de la forma

$$\lambda \sum [(1 - \alpha) |b_p| + \alpha |b_p|^2]$$



De esta forma cuando  $\lambda \neq 0$  se estaría en presencia del contexto de penalización y selección de variables. Contrariamente, si el valor de  $\lambda$  es cero, la estimación es la clásica OLS. Por otro parte si  $\alpha=1$ , de manera que sólo se penaliza cuadráticamente la regresión resultante es de tipo RIDGE, y si  $\alpha=0$  entonces la estimación pasa a ser en el marco de regresores LASSO (Varian, 2014).

La utilización de LASSO, se explica en que al no poseer el término cuadrática, las soluciones son de esquina, haciendo coeficientes iguales a cero (y por ende seleccionando al resto de las variables) (James et al., 2013)

Este método penaliza por cantidad de variables utilizadas, reduciendo a cero los coeficientes de las que menos contribuyen a la explicación del modelo. Según el parámetro de penalización, los coeficientes van reduciéndose hasta hacerse cero. El resultado de los estimadores que se muestran, es el que resulta del  $\lambda$  óptimo dado por la minimización del error medio cuadrático, la cual es realizada mediante la técnica de cross-validation.

Un aspecto a resaltar es que mediante LASSO se pierden los estadísticos clásicos, dada la imposibilidad de garantizar la distribución que se suele ser asumida con p-valores. No puede evaluarse la existencia de significatividad. Sin embargo, a diferencia de las tradicionales estimaciones realizadas mediante OLS, se obtiene una varianza inferior (Oladunni et al., 2017).

### **3.2 Classification and Regression Tree (CART)**

Frente a la utilización de predictores globales, donde una única fórmula predictiva se utiliza para todo el conjunto de datos, la existencia de diversas características que interactúan de manera no lineal, hacen que la realización de un único modelo, sea difícil de ser estimada.

De esta manera, una opción para predecir la respuesta de la variable explicativa es recurrir al método de CART. El mismo hace uso de un árbol de decisión binario, de manera tal que cada nodo del mismo, testeé alguna de las variables explicativas. Dependiendo, del resultado, el próximo paso, consiste en estudiar alguna de las subdivisiones generadas por la rama del árbol. El paso final de estas estimaciones, se produce al final del árbol (hoja), donde se realiza una predicción., la cual promedia todos los datos agregados para esta partición.

No obstante la división de nodos, debe implementar algún criterio de separación válido, siendo uno de los más utilizados el que consiste en la suma de errores cuadrados de un árbol T de forma que

$$S = \sum_{c \in \text{hojas}(T)} \sum_{i \in c} (y_i - m_c)^2$$

Siendo  $m_c = \frac{1}{n_c} \sum_{i \in c} y_i$ , la predicción para  $y_i$  para la hoja  $c$ , la expresión anterior puede reescribirse como

$$S = \sum_{c \in \text{hojas}(T)} n_c V_c$$

Donde  $V_c$  es la varianza dentro de la hoja  $c$ , lo cual minimizaría a  $S$ .

#### 4. Análisis de resultados

Para el caso específico de LASSO, mediante la utilización del paquete *glmnet*, en R (Friedman et al, 2008), se procedió a la realización de estimaciones. Se detalla en la tabla 2, aquellas variables cuyo coeficiente resultó distinto a cero. Luego de hallar el valor lambda óptimo (coeficiente de penalización), el número de variables se reduce a 39.

Hay muchos casos donde no sorprende que una variable sea importante para determinar el valor de un auto usado, es el caso del kilometraje, y la antigüedad, cuya relación respecto del precio, es consistente con lo esperado. Automóviles con mayor cantidad de kilómetros recorridos como así también una mayor cantidad de años desde su fabricación, tendrán un precio menor.

En este sentido, también debe destacarse la relevancia de que el coche sea de tipo utilitario. Automóviles de este tipo, están expuestos a un mayor desgaste, dado que el uso del mismo se refiere a actividades de carga.

Un interesante resultado, se observa al momento de evaluar si el auto es de único dueño. Desde la perspectiva del consumidor, el signo positivo, hallaría su razón de ser en la percepción y confianza generadas a causa de que el propietario, conocería totalmente el vehículo en venta. Empero, también podría indicar que aquellos vendedores profesionales y los cuales no son los primeros poseedores, proceden a colocar en venta los vehículos con una cotización menor, a fin de poder desprenderse del vehículo en el menor lapso temporal posible.

Por otra parte, algunas características que dentro de un auto nuevo incrementan su valor simplemente por poseerlas, también adquieren relevancia para el caso de los coches de segunda mano. Es el caso de la dirección automática, reproductor de DVD, GPS, sensor de estacionamiento y lluvia, vidrios blindados, potencia y control de tracción.

No obstante, en algunas variables específicas se encuentran características llamativas. El más interesante corresponde a la existencia de distintos colores en este sub grupo de variables seleccionadas. En el caso de aquellos automóviles de color amarillo, el coeficiente puede explicarse por el valor de modelos disponible en esta tonalidad como así también que estos constituyen una rareza, producto de las políticas comerciales de las compañías automotrices<sup>3</sup>. Situación diametralmente opuesta ocurre con los vehículos cuya coloración es verde o azul, cuya oferta es superior.

A su vez, se resalta el valor negativo del parámetro asociado a la cantidad de puertas. Una potencial explicación puede hallarse en que aquellos autos de 5 puertas son los más convencionales comúnmente de uso familiar, mientras que aquellos con menor cantidad se encuentran asociados a modelos de tipo deportivo los cuales poseen un precio superior al promedio.

Puede considerarse también que los consumidores evalúan el precio del combustible como una cuestión de importancia al momento de realizar su compra: existe un impacto positivo en el precio si el automóvil es de tipo diésel, teniendo en cuenta que este último resulta más barato que la nafta convencional.

Un regresor de relevancia es el correspondiente a si el auto es de alta gama. El mismo, es el más elevado de toda la estimación, lo cual estaría indicando el fuerte impacto positivo de en el precio de venta de los automóviles con esta característica, lo cual resulta esperable. Vehículos de este tipo son considerados exclusivos, y con una serie de atributos distintivos, respecto del resto de automóviles disponibles en el mercado.

También debe indicarse la importancia de la carrocería. La cotización final de un vehículo, resulta afectada positivamente en el caso de los pick-ups, dado que los mismos podrían ser asociados a una combinación de confort, calidad y en simultáneo, ser de carga.

Otros determinantes de importancia que deben mencionarse son la existencia de Air-Bag y doble tracción, cuestiones que incrementan el valor final de venta. Debe resaltarse que ambos elementos tienen un costo de colocación y reparación elevado.

**Tabla 2. Variables estimadas mediante LASSO.**

<b>Variable</b>	<b>Coeficiente</b>	<b>Variable</b>	<b>Coeficiente</b>
Intercepto	-10028340.00	DVD	13199.75

<sup>3</sup> Según una investigación realizada por la compañía iSeeCars.com mediante el relevamiento 1.6 millones de autos usados, durante los años 2015 y 2016, se destaca que los vehículos de color amarillo y verde representan el 1,5% del total de automóviles en venta (Lee, 2017)

Dirección	-1347.98	Sensor de estacionamiento	12281.26
Único dueño	12229.70	Sensor de lluvia	31591.00
Combustible diésel	4141.54	Airbag de cortina	50039.50
Puertas	-8274.69	Asientos eléctricos	45982.80
KM	-0.34	Blindado	30604.47
Transmisión	18268.22	Doble tracción	52839.37
Paragolpes pintados	-5897.17	GPS	43412.56
AM/FM	-4829.99	Pasacasete	-42468.86
Entrada auxiliar	264.36	Importado	28702.05
Aire Acondicionado	-24418.31	Alta gama	79657.04
Asiento trasero rebatible	-4247.72	Utilitario	-982.54
Tercera luz de freno	-7751.30	Pickup	31423.92
Apoyacabeza en asiento trasero	-578.44	década	2298.08
Potencia	97.87	MERCOSUR	-40781.22
Caja de CD	-7759.09	Antigüedad	-29.97
Tarjeta SD	48.47	Real kilómetros/Años.	0.19
Control de tracción	4503.73	Verde	-680.82
Silla de bebé	777.69	Azul	-9222.21
		Amarillo	72824.69

Fuente: elaboración propia

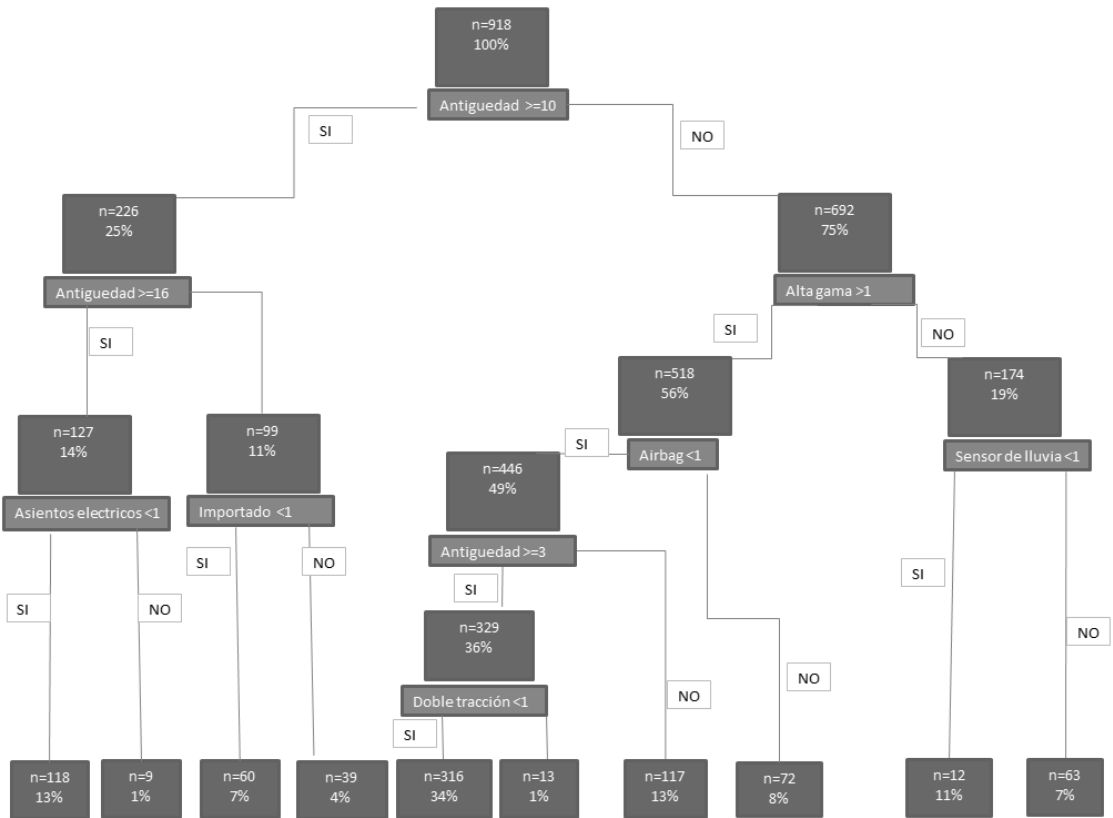
Por el contrario, el coeficiente que más reduce la cotización de un coche usado, es el que releva si este último posee pasacasete. Dicha cualidad se encuentra asociada a una tecnología en desuso. A la par, puede reflejar la falta de mantenimiento de aquellos autos con sistema reproductor de música obsoleto, es decir, el propietario no ha tenido intención alguna de mejorar esta característica básica del automóvil. Contrariamente, los vehículos que disponen de ranura para introducir tarjetas SD, se corresponden con un mayor valor de venta.

Asimismo, es valorizado negativamente que el país de fabricación pertenezca al MERCOSUR, lo cual puede estar asociado a las exenciones impositivas existentes para los integrantes de este bloque, como así también la ausencia de fabricación de modelos con un precio de venta elevado por parte de esos países. Contrariamente, aquellos automóviles importados de otras naciones, dado el arancelamiento que posee este sector, poseen un valor final de comercialización superior.

Al momento de aplicar CART, se hizo uso del paquete *rpart* de R (Therneau et al., 2010). De esta forma, tal como se ilustra en el gráfico 1, puede observarse a primera vista el fuerte impacto de la antigüedad, en el árbol de decisión. Además esto último posibilita detectar una cuestión interesante: la mayoría de automóviles en venta, específicamente el 75%, tienen menos de una década desde el momento de su fabricación. De la misma forma, sobresale el sustancial porcentaje, un 34%, de aquellos que son de alta gama, con airbag y doble tracción. Esta

situación evidenciaría un alto segmento del mercado, que ofrece a la venta autos considerados costosos.

**Gráfico 1. Variables representadas mediante CART**



Fuente: Elaboración propia.

Asimismo, mediante este tipo de estimación puede señalarse que como algunas variables dummies, que mediante LASSO, tenían un fuerte impacto sobre el precio, adquieren relevancia de forma tal que son separadas en ramas. Concretamente es la situación de importados, asientos eléctricos, sensores de lluvia, air-bag y sensor de lluvia. De esta manera, a excepción de la cantidad de años, el resto de regresores es de tipo binario.

Sin embargo, resulta necesario destacar que para el caso de CART, sorprende la ausencia de la variable referente al color amarillo, la cual era señalada como uno de los coeficientes de mayor impacto positivo.

También debe mencionarse, que este tipo de estimaciones no releva algunos atributos que esperaban ser considerados. Un ejemplo de esto es el kilometraje, como así también la presencia de un único dueño.

## 5. Conclusiones

El análisis realizado en este trabajo consistió en la aplicación de métodos de Machine Learning, recurriendo concretamente a estimaciones econométricas de tipo LASSO y CART, a fin de evaluar el impacto de los diversos determinantes considerados en el precio de los automóviles usados dentro de Argentina.

En este sentido, LASSO tiene como principal ventaja reducir el número de variables explicativas, mientras que CART, posibilita analizar regresiones estrictamente no lineales, permitiendo un interesante análisis gráfico.

Ambos métodos deben considerarse no como excluyentes sino más precisamente como complementarios. Se resalta para los mismos, como aspecto común, la fuerte afectación negativa en el valor final dado por la antigüedad del automóvil.

Por otra parte, también se observa que específicamente los autos de alta gama y con pocos años de fabricación, parecen ser el segmento concreto de coches usados que poseen una mayor valoración dentro del mercado.

Se enfatiza la ausencia de trabajos que recurran a ML destinados a la evaluación de modelos de precios hedónicos, concretamente para el ámbito nacional. Esta línea de investigación tiene aún un incipiente campo de investigación en el cual la multidisciplinariedad entre las ciencias de la computación y la economía merecen desarrollar.

Se postula como futuras líneas de investigación el estudio de aquellas variables vinculadas directamente con los diferenciales geográficos, como así también la depreciación real como determinantes que influyen sobre el precio final de los automóviles usados.

## Referencias

- Akay, E., Bölükbaşı, Ö. y Bekar E. (2018): “Robust and Resistant Estimations of Hedonic Prices for Second Hand Cars: An Application to the Istanbul Car Market”, *International Journal of Economics and Financial Issues*, 8(1):39-47.
- Akerlof, G. (1970), “The Market for "Lemons": Quality Uncertainty and the Market Mechanism”, *The Quarterly Journal of Economics*, (84) 3: 488-500. Recuperado

de

[https://www.sas.upenn.edu/~hfang/teaching/socialinsurance/readings/fudan\\_hsb/c/Akerlof70\(2.1\).pdf](https://www.sas.upenn.edu/~hfang/teaching/socialinsurance/readings/fudan_hsb/c/Akerlof70(2.1).pdf)

Baltas, G. y C. Saridakis (2009), “Brand-name effects, segment differences, and product characteristics: an integrated model of the car market”, *Journal of Product & Brand Management* 18/2: 143–151.  
<http://dx.doi.org/10.1108/10610420910949040>

Boyd, J. y Mellman R. (1980), “The Effect of Fuel Economy Standard on the U.S. Automotive Market: Am Hedonic Demand Analysis”, *Transportation Research Part A: General*, (14) 5: 367-380. [https://doi.org/10.1016/0191-2607\(80\)90055-2](https://doi.org/10.1016/0191-2607(80)90055-2)

Court, A.T. (1939), “Hedonic price indexes with automotive examples”. En *The Dynamics of Automobile Demand*. Nueva York: General Motors Corporation, 99-119.

Cowling, K. y Cubbin, J. (1972): “Hedonic Price Indexes for United Kingdom Cars”. *The Economic Journal*, 82 (327): 963-978.

Cropper, M., L. Deck, y K. McConnell (1988), “On the Choice of Functional Form for Hedonic Price Functions”, *Review of Economics and Statistics* 70: 668-675.

Desmet P. y Hendoui F. (2000), “La relation prix-qualité dans l’automobile: comparaison de méthodes d’estimation des prix hédoniques”, *Revue Française du Marketing*, vol. 179-180: 167-179.

Dvir E. Strasser G. (2018): “Does marketing widen borders? Cross-country price dispersion in the European car market”, *Journal of International Economics*, (112): 134-149.  
<https://doi.org/10.1016/j.jinteco.2018.02.008> .

Erdem, C. y I. Şentürk (2009), “A Hedonic Analysis of Used Car Prices in Turkey”, *International Journal of Economic Perspectives* 3 (2): 141-149.

- Fan, Q. y Rubin J. (2010): “Two-Stage Hedonic Price Model for Light-Duty Vehicles”, *Transportation Research Record: Journal of the Transportation Research Board*, 1 (2157), 119-128.
- Friedman, J., Trevor, H. y Tibshirani, R. (2008), Regularization Paths for Generalized Linear Models via Coordinate Descent, *Journal of Statistical Software*, 33(1): 1-22.
- Gilbert, S (2013), “Applying the Hedonic Method”, *National Institute of Standards and Technology* Technical Note 1811, <http://dx.doi.org/10.6028/NIST.TN.1811> .
- González, E., Cárcaba A. y Ventura J. (2015), “How car dealers adjust prices to reach the product efficiency frontier in the Spanish automobile markets”, *Omega*, (5): 38-48. <https://doi.org/10.1016/j.omega.2014.08.005>
- Griliches, Z. (1961), “Hedonic Price Indexes for Automobiles: An Econometric of Quality Change”. En D.S. Brady; E.F. Denison; I.B. Kravis; P.J. McCarthy; A. Rees; R. Ruggles; B.C. Swerling y G.J. Stigler (eds.): *The Price Statistics of the Federal Government*. Nueva York: National Bureau of Economic Research, 137-196.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). New York: Springer
- Kihm, A. y C. Vance (2014), “The Determinants of Equity Transmission between the New and Used Car Markets – A Hedonic Analysis”, Ruhr-Universität Bochum (RUB) Economic Paper No. 521. <http://dx.doi.org/10.4419/86788596>
- Kiso , T. (2015) “Pricing Automobile Fuel Economy : A New Hedonic Approach “ *Discussion Paper in Economics* , 15 (19) , University of Aberdeen Business School , Aberdeen.
- Knoor, K. y Jan S. (2017): “Vehicle Price Prediction System using Machine Learning Techniques”, *International Journal of Computer Applications*, 167 (9):27-31.
- Lancaster, K.J. (1966): “A New Approach to Consumer Theory”, *The Journal of Political Economy*, 74 (2): 132-157.
- Ng, S. (2017). Opportunities and challenges: Lessons from analyzing terabytes of scanner data (No. w23673). National Bureau of Economic Research.
- Oladunni, T., Sharad, S., y Tiwang R. (2017) “A Spatio – Temporal Hedonic House Regression Model” *16th IEEE International Conference on Machine Learning and Applications*: 607-612.



- Prieto, C., B. Caemmerer, y G. Baltas (2015), “Using a hedonic price model to test prospect theory assertions: The asymmetrical and nonlinear effect of reliability on used car prices”, *Journal of Retailing and Consumer Services*, 22: 206–212.
- Pudaruth, S. (2014): “Predicting the Price of Used Cars using Machine Learning Techniques”, *International Journal of Information & Computation Technology*, 4 (7): 753-764.
- Ramírez Muñoz de Toro, G., Uriarte, J.I., Delbianco, F. y Larrosa, JMC (2017): “Un modelo hedónico de precios en línea de automóviles usados en Argentina”, *Revista de métodos cuantitativos para la economía y la empresa*, (24): 25-53.
- Rosen, S. (1974), “Hedonic Prices and Implicit markets: Product Differentiation in Pure Competition”. *Journal of Political Economy*, 82(1): 34-55.
- Sallee, J., West, S. y Fan, W. (2016), “Do consumers recognize the value of fuel economy? Evidence from used car prices and gasoline price fluctuations”, *Journal of Public Economics*, (135): 61-73. <https://doi.org/10.1016/j.jpubeco.2016.01.003>
- Therneau, T. M., Atkinson, B., y Ripley, M. B. (2010). The rpart package.
- Varian, H. R. (2014). Big data: New tricks for econometrics, *Journal of Economic Perspectives*, 28(2): 3-28.
- Yang M. (2018), *Essays on Machine Learning and Hedonic Models* (tesis doctoral), University of Washington, Estados Unidos.